US008990125B2

(12) United States Patent
Raichelgauz et al.

(10) **Patent No.: US 8,990,125 B2**
(45) **Date of Patent: Mar. 24, 2015**

(54) **SIGNATURE GENERATION FOR MULTIMEDIA DEEP-CONTENT-CLASSIFICATION BY A LARGE-SCALE MATCHING SYSTEM AND METHOD THEREOF**

(71) Applicant: **Cortica, Ltd.**, Tital Carmel (IL)

(72) Inventors: **Igal Raichelgauz**, Herzelia (IL); **Karina Ordinaev**, Haifa (IL); **Yehoshua Y. Zeevi**, Haifa (IL)

(73) Assignee: **Cortica, Ltd.**, Ramat Gan (IL)

(*) Notice: Subject to any disclaimer, the term of this patent is extended or adjusted under 35 U.S.C. 154(b) by 181 days.

(21) Appl. No.: **13/682,132**

(22) Filed: **Nov. 20, 2012**

(65) **Prior Publication Data**

US 2013/0080433 A1 Mar. 28, 2013

Related U.S. Application Data

(63) Continuation of application No. 12/195,863, filed on Aug. 21, 2008, now Pat. No. 8,326,775, and a continuation-in-part of application No. 12/084,150, filed as application No. PCT/IL2006/001235 on Oct. 26, 2006, now Pat. No. 8,655,801.

(30) **Foreign Application Priority Data**

Oct. 26, 2005	(IL)	171577
Jan. 29, 2006	(IL)	173409
Aug. 21, 2007	(IL)	185414

(51) **Int. Cl.**
G06G 7/00 (2006.01)
G06E 1/00 (2006.01)
(Continued)

(52) **U.S. Cl.**
CPC ***G06F 17/30595*** (2013.01); ***G06F 17/30743*** (2013.01); ***G06F 17/30746*** (2013.01);
(Continued)

(58) **Field of Classification Search**
None
See application file for complete search history.

(56) **References Cited**

U.S. PATENT DOCUMENTS

6,243,375	B1	6/2001	Speicher
6,611,628	B1	8/2003	Sekiguchi et al.

(Continued)

FOREIGN PATENT DOCUMENTS

WO	02/31764	4/2002
WO	2007/0049282	5/2007

OTHER PUBLICATIONS

Role of Spatiotemporal Oriented Energy Features for Robust Visual Tracking in Video Surveillance Emami, A. ; Dadgostar, F. ; Bigdeli, A. ; Lovell, B.C. Advanced Video and Signal-Based Surveillance (AVSS), 2012 IEEE Ninth International Conference on DOI: 10.1109/AVSS.2012.64 Publication Year: 2012 , pp. 349-354.*

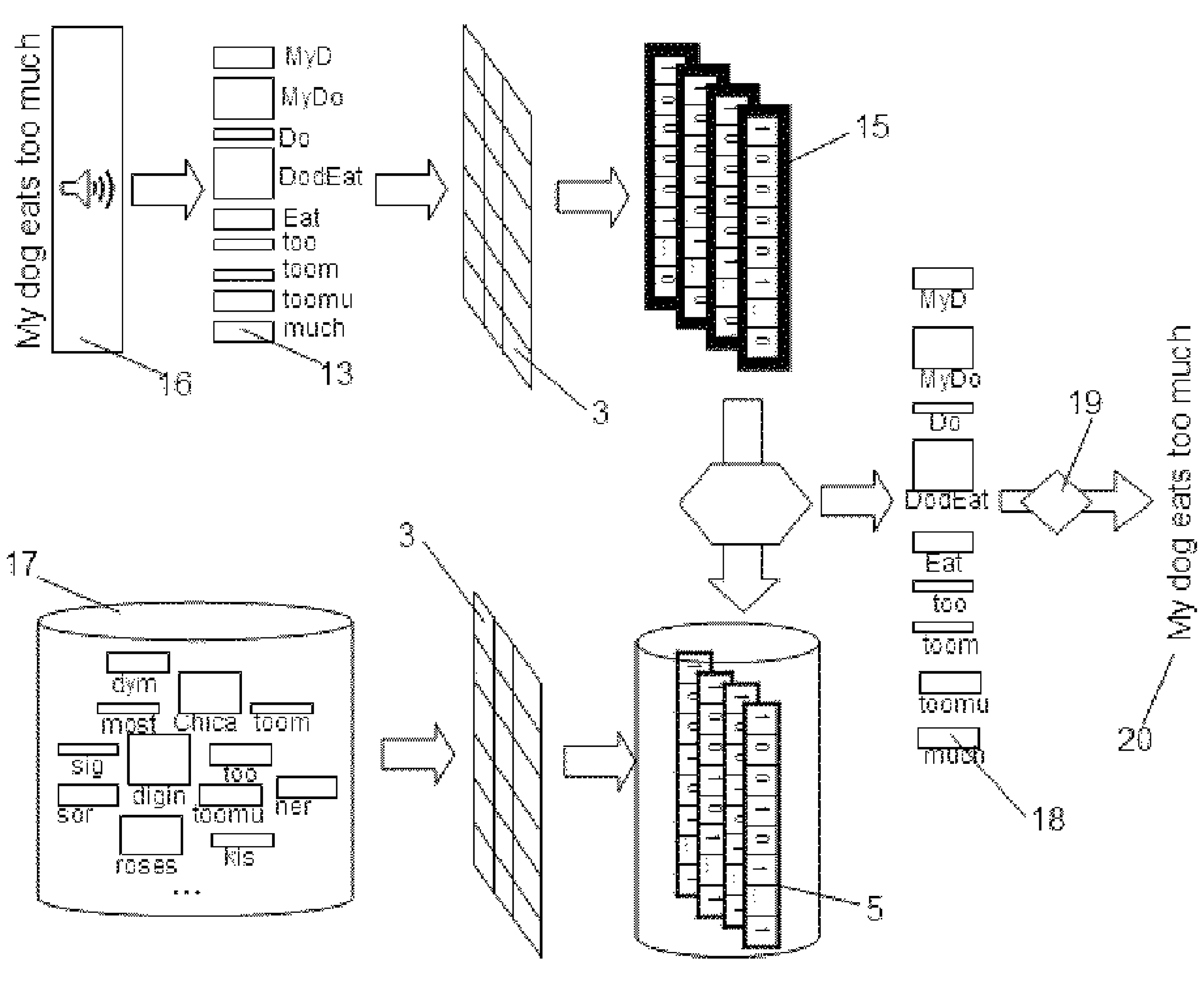
(Continued)

Primary Examiner — Michael B Holmes
(74) *Attorney, Agent, or Firm* — M&B IP Analysts, LLC

(57) **ABSTRACT**

Content-based clustering, recognition, classification and search of high volumes of multimedia data in real-time. The embodiments disclosed herein are dedicated to real-time fast generation of signatures to high-volume of multimedia content-segments, based on relevant audio and visual signals, and to scalable matching of signatures of high-volume database of content-segments' signatures. The embodiments disclosed herein can be implemented in any applications which involve large-scale content-based clustering, recognition and classification of multimedia data, such as, content-tracking, video filtering, multimedia taxonomy generation, video fingerprinting, speech-to-text, audio classification, object recognition, video search and any other application requiring content-based signatures generation and matching for large content volumes such as, web and other large-scale databases.

13 Claims, 6 Drawing Sheets



(51) **Int. Cl.**
G06E 3/00 (2006.01)
G06F 15/00 (2006.01)
G06N 1/00 (2006.01)
G06F 17/30 (2006.01)
G06K 9/00 (2006.01)
G10L 15/26 (2006.01)
G10L 15/32 (2013.01)

(52) **U.S. Cl.**
CPC ***G06F17/30787*** (2013.01); ***G06F 17/30796*** (2013.01); ***G06F 17/30799*** (2013.01); ***G06K 9/00744*** (2013.01); ***G10L 15/265*** (2013.01); *G10L 15/32* (2013.01)
USPC **706/10**

(56) **References Cited**

U.S. PATENT DOCUMENTS

6,754,435	B2	6/2004	Kim	
6,819,797	B1	11/2004	Smith et al.	
6,901,207	B1	5/2005	Watkins	
7,013,051	B2	3/2006	Sekiguchi et al.	
7,260,564	B1	8/2007	Lynn et al.	
7,302,117	B2	11/2007	Sekiguchi et al.	
7,313,805	B1	12/2007	Rosin et al.	
8,112,376	B2 *	2/2012	Raichelgauz et al.	706/46
8,266,185	B2 *	9/2012	Raichelgauz et al.	707/803
8,312,031	B2 *	11/2012	Raichelgauz et al.	707/756
8,326,775	B2 *	12/2012	Raichelgauz et al.	706/10
8,386,400	B2 *	2/2013	Raichelgauz et al.	706/12
8,655,801	B2 *	2/2014	Raichelgauz et al.	706/12
8,799,195	B2 *	8/2014	Raichelgauz et al.	706/12
8,799,196	B2 *	8/2014	Raichelquaz et al.	706/12
8,818,916	B2 *	8/2014	Raichelgauz et al.	706/10
8,868,619	B2 *	10/2014	Raichelgauz et al.	707/803
8,880,539	B2 *	11/2014	Raichelgauz et al.	707/756
8,880,566	B2 *	11/2014	Raichelgauz et al.	707/803
2001/0019633	A1	9/2001	Tenze et al.	
2003/0041047	A1	2/2003	Chang et al.	
2004/0153426	A1	8/2004	Nugent	
2005/0177372	A1	8/2005	Wang et al.	
2006/0253423	A1	11/2006	McLane et al.	
2007/0074147	A1	3/2007	Wold	
2007/0130159	A1	6/2007	Gulli et al.	
2007/0244902	A1	10/2007	Seide et al.	
2007/0253594	A1	11/2007	Lu et al.	
2008/0072256	A1	3/2008	Boicey et al.	

OTHER PUBLICATIONS

Emotional speech characterization based on multi-features fusion for face-to-face interaction Mandhaoui, A. ; Ringeval, F. ; Chetouani, M. Signals, Circuits and Systems (SCS), 2009 3rd International Conference on DOI: 10.1109/ICSCS.2009.5412691 Publication Year: 2009 , pp. 1-6.*

Automatic Diagnosis of Defects of Rolling Element Bearings Based on Computational Intelligence Techniques Cococcioni, M. ; Lazzerini, B. ; Volpi, S.L. Intelligent Systems Design and Applications, 2009. ISDA '09. Ninth International Conference on DOI: 10.1109/ISDA.2009.240 Publication Year: 2009 , pp. 970-975.*

Real time speaker localization and detection system for camera steering in multiparticipant videoconferencing environments Marti, Amparo ; Cobos, Maximo ; Lopez, Jose J. Acoustics, Speech and Signal Processing (ICASSP), 2011 IEEE International Conference on DOI: 10.1109/ICASSP.2011.5947015 PublicationYear: 2011 , pp. 2592-2595.*

Verstraeten et al., "Isolated word recognition with the Liquid State Machine: a case study"; Department of Electronics and Information Systems, Ghent University, Sint-Pietersnieuwstraat 41, 9000 Gent, Belgium, Available online Jul. 14, 2005.

Zhou et al., "Medical Diagnosis With C4.5 Rule Preceded by Artificial Neural Network Ensemble"; IEEE Transactions on Information Technology in Biomedicine, vol. 7, Issue: 1, pp. 37-42, Date of Publication: Mar. 2003.

Cernansky et al., "Feed-forward Echo State Networks"; Proceedings of International Joint Conference on Neural Networks, Montreal, Canada, Jul. 31-Aug. 4, 2005.

Lyon, Richard F.; "Computational Models of Neural Auditory Processing"; IEEE International Conference on Acoustics, Speech, and Signal Processing, ICASSP '84, Date of Conference: Mar. 1984, vol. 9, pp. 41-44.

Freisleben et al., "Recognition of Fractal Images Using a Neural Network", Lecture Notes in Computer Science, 1993, vol. 6861, 1993, pp. 631-637.

Ware et al., "Locating and Identifying Components in a Robot's Workspace using a Hybrid Computer Architecture"; Proceedings of the 1995 IEEE International Symposium on Intelligent Control, Aug. 27-29, 1995, pp. 139-144.

Zhou et al., "Ensembling neural networks: Many could be better than all"; National Laboratory for Novel Software Technology, Nanjing Unviersirty, Hankou Road 22, Nanjing 210093, PR China; Received Nov. 16, 2001, Available online Mar. 12, 2002.

Fathy et al., "A Parallel Design and Implementation for Backpropagation Neural Network Using NIMD Architecture", 8th Mediterranean Electrotechnical Corsfe rersce, 19'96. MELECON '96, Date of Conference: May 13-16, 1996, vol. 3, pp. 1472-1475.

Howlett et al., "A Multi-Computer Neural Network Architecture in a Virtual Sensor System Application", International Journal of Knowledge-based Intelligent Engineering Systems, 4 (2). pp. 86-93, 133N 1327-2314; first submitted Nov. 30, 1999; revised version submitted Mar. 10, 2000.

Ortiz-Boyer et al., "CIXL2: A Crossover Operator for Evolutionary Algorithms Based on Population Features", Journal of Artificial Intelligence Research 24 (2005) 1-48 Submitted Nov. 2004; published Jul. 2005.

IPO Examination Report under Section 18(3) for corresponding UK application No. GB1001219.3, dated May 30, 2012.

IPO Examination Report under Section 18(3) for corresponding UK application No. GB1001219.3, dated Sep. 12, 2011.

Lin, C.; Chang, S.: "Generating Robust Digital Signature for Image/Video Authentication", Multimedia and Security Workshop at ACM Mutlimedia '98; Bristol, U.K., Sep. 1998; pp. 49-54.

Iwamoto, K.; Kasutani, E.; Yamada, A.: "Image Signature Robust to Caption Superimposition for Video Sequence Identification"; 2006 IEEE International Conference on Image Processing; pp. 3185-3188, Oct. 8-11, 2006; doi: 10.1109/ICIP.2006.313046.

Maass, W. et al.: "Computational Models for Generic Cortical Microcircuits", Institute for Theoretical Computer Science, Technische Universitaet Graz, Graz, Austria, published Jun. 10, 2003.

International Search Report for the corresponding International Patent Application PCT/IL2006/001235; Date of Mailing: Nov. 2, 2008.

Raichelgauz, I. et al.: "Co-evolutionary Learning in Liquid Architectures", Lecture Notes in Computer Science, [Online] vol. 3512, Jun. 21, 2005, pp. 241-248, XP019010280 Springer Berlin / Heidelberg ISSN: 1611-3349 ISBN: 978-3-540-26208-4.

Jaeger, H.: "The "echo state" approach to analysing and training recurrent neural networks", GMD Report, No. 148, 2001, pp. 1-43, XP002466251. German National Research Center for Information Technology.

Verstraeten et al.: "Isolated word recognition with the Liquid State Machine: a case study", Information Processing Letters, Amsterdam, NL, vol. 95, No. 6, Sep. 30, 2005, pp. 521-528, XP005028093 ISSN: 0020-0190.

Zeevi, Y. et al.: "Natural Signal Classification by Neural Cliques and Phase-Locked Attractors", IEEE World Congress on Computational Intelligence, IJCNN2006, Vancouver, Canada, Jul. 2006, XP002466252.

Natsclager, T. et al.: "The "liquid computer": A novel strategy for real-time computing on time series", Special Issue on Foundations of Information Processing of Telematik, vol. 8, No. 1, 2002, pp. 39-43, XP002466253.

Morad, T.Y. et al.: "Performance, Power Efficiency and Scalability of Asymmetric Cluster Chip Multiprocessors", Computer Architecture Letters, vol. 4, Jul. 4, 2005, pp. 1-4, XP002466254.

Burgsteiner et al.: "Movement Prediction From Real-World Images Using a Liquid State Machine", Innovations in Applied Artificial

(56) **References Cited**

OTHER PUBLICATIONS

Intelligence Lecture Notes in Computer Science, Lecture Notes in Artificial Intelligence, LNCS, Springer-Verlag, BE, vol. 3533, Jun. 2005, pp. 121-130.

International Search Authority: "Written Opinion of the International Searching Authority" (PCT Rule 43bis.1) including International Search Report for International Patent Application No. PCT/US2008/073852; Date of Mailing: Jan. 28, 2009.

Xian-Sheng Hua et al.: "Robust Video Signature Based on Ordinal Measure" In: 2004 International Conference on Image Processing, ICIP '04; Microsoft Research Asia, Beijing, China; published Oct. 24-27, 2004, pp. 685-688.

International Search Authority: International Preliminary Report on Patentability (Chapter I of the Patent Cooperation Treaty) including "Written Opinion of the International Searching Authority" (PCT Rule 43bis. 1) for the corresponding International Patent Application No. PCT/IL2006/001235; Date of Issuance: Jul. 28, 2009.

IPO Examination Report under Section 18(3) for corresponding UK application No. GB 1001219.3, dated Aug. 10, 2012.

* cited by examiner

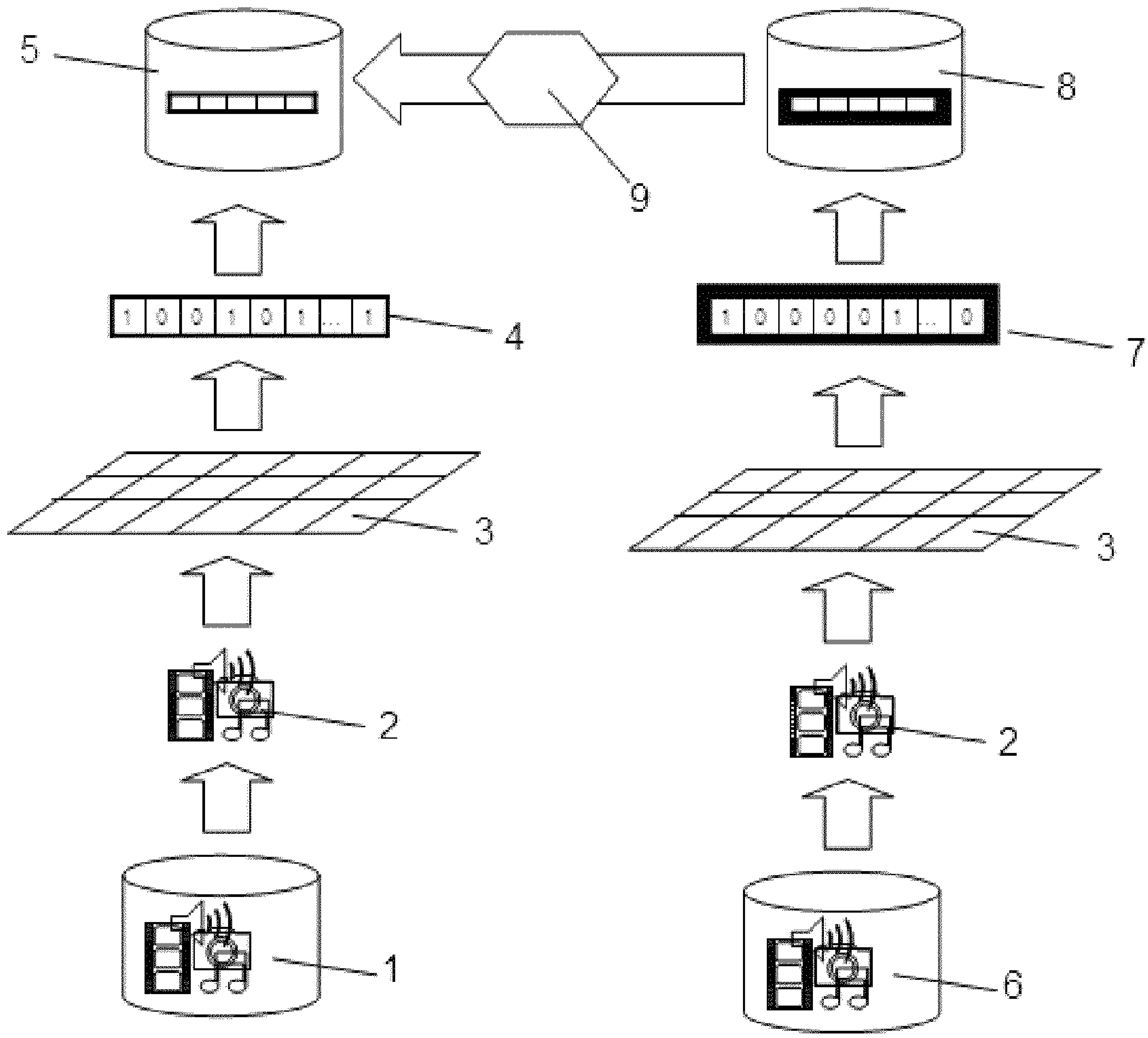


FIG. 1

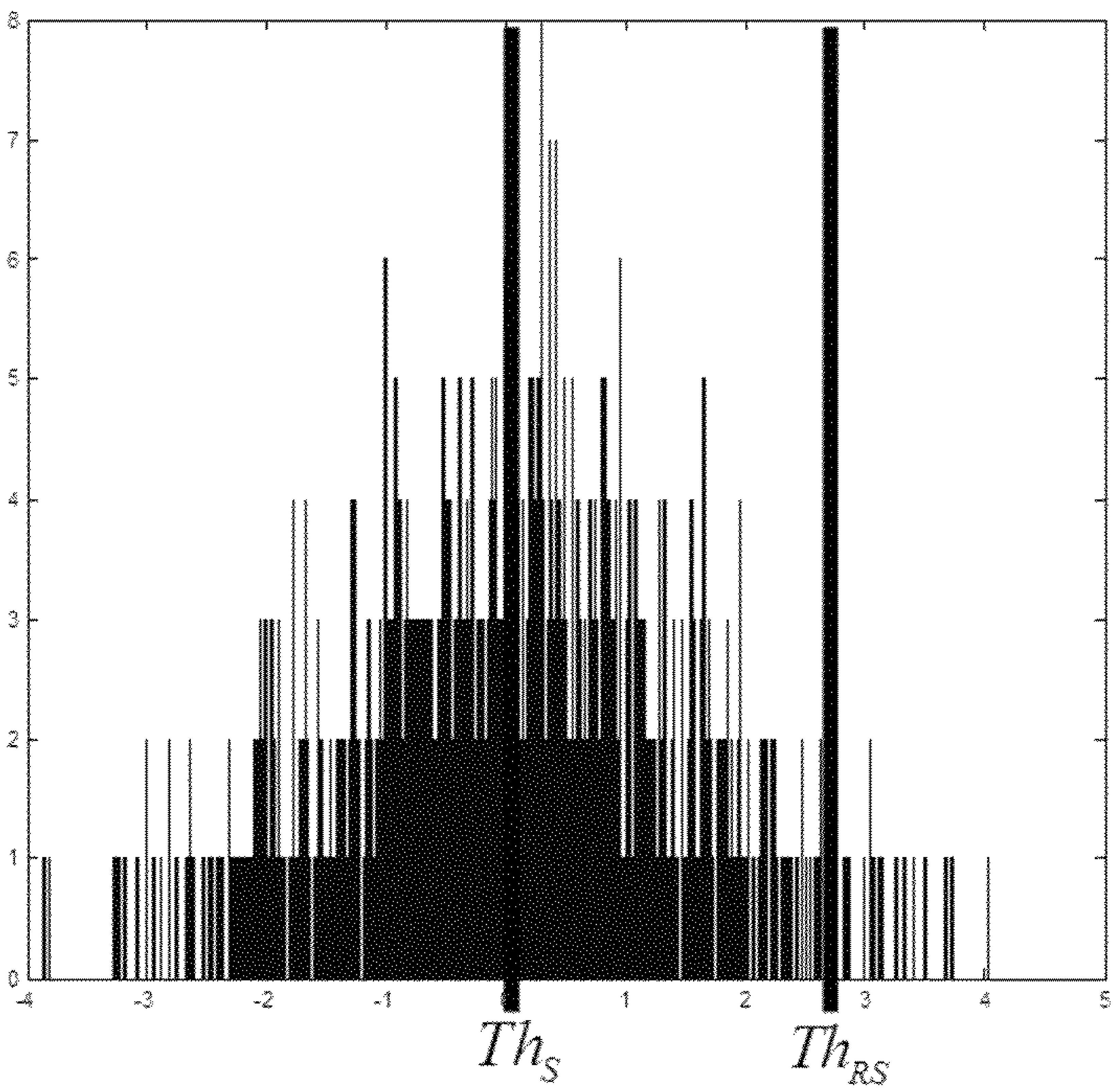


FIG. 2

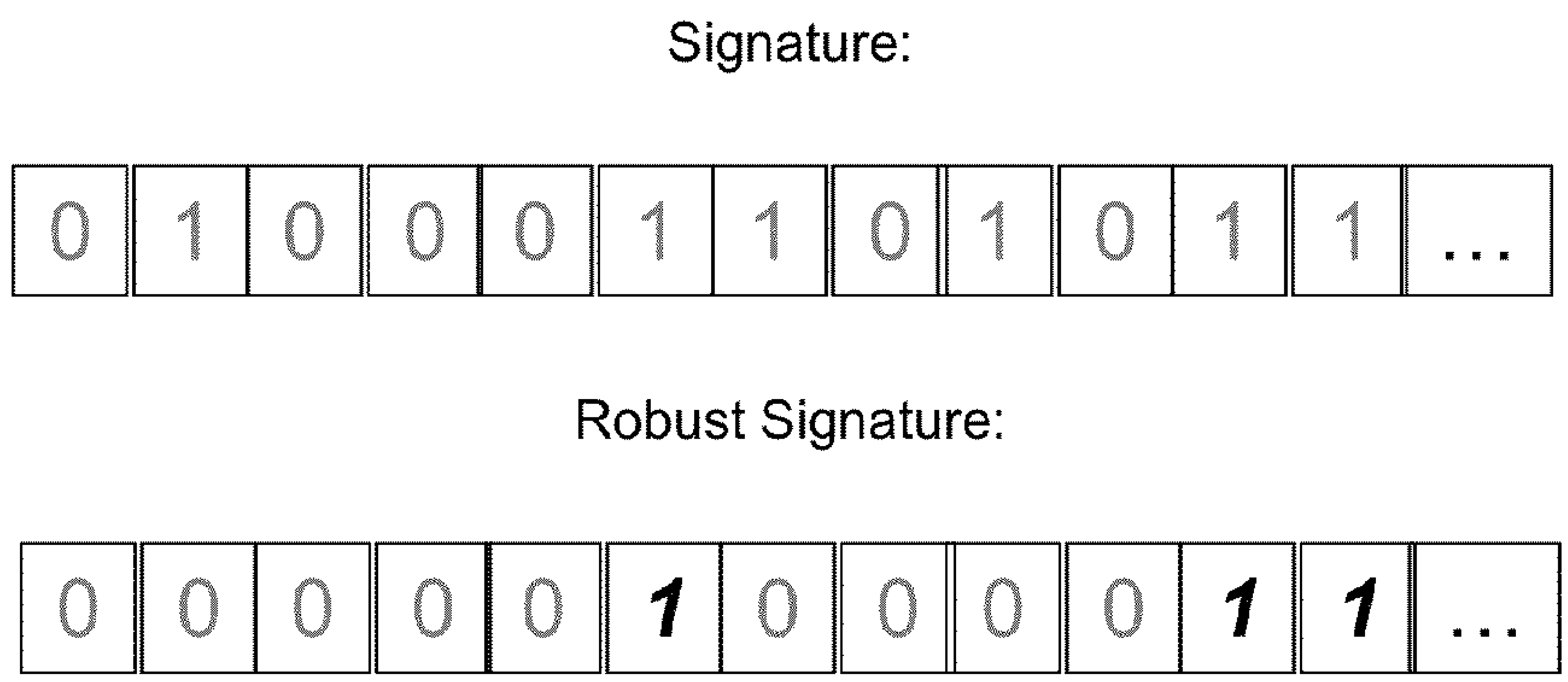

Signature:
0 1 0 0 0 1 1 0 1 0 1 1 …
Robust Signature:
0 0 0 0 0 1 0 0 0 0 1 1 …

FIG. 3

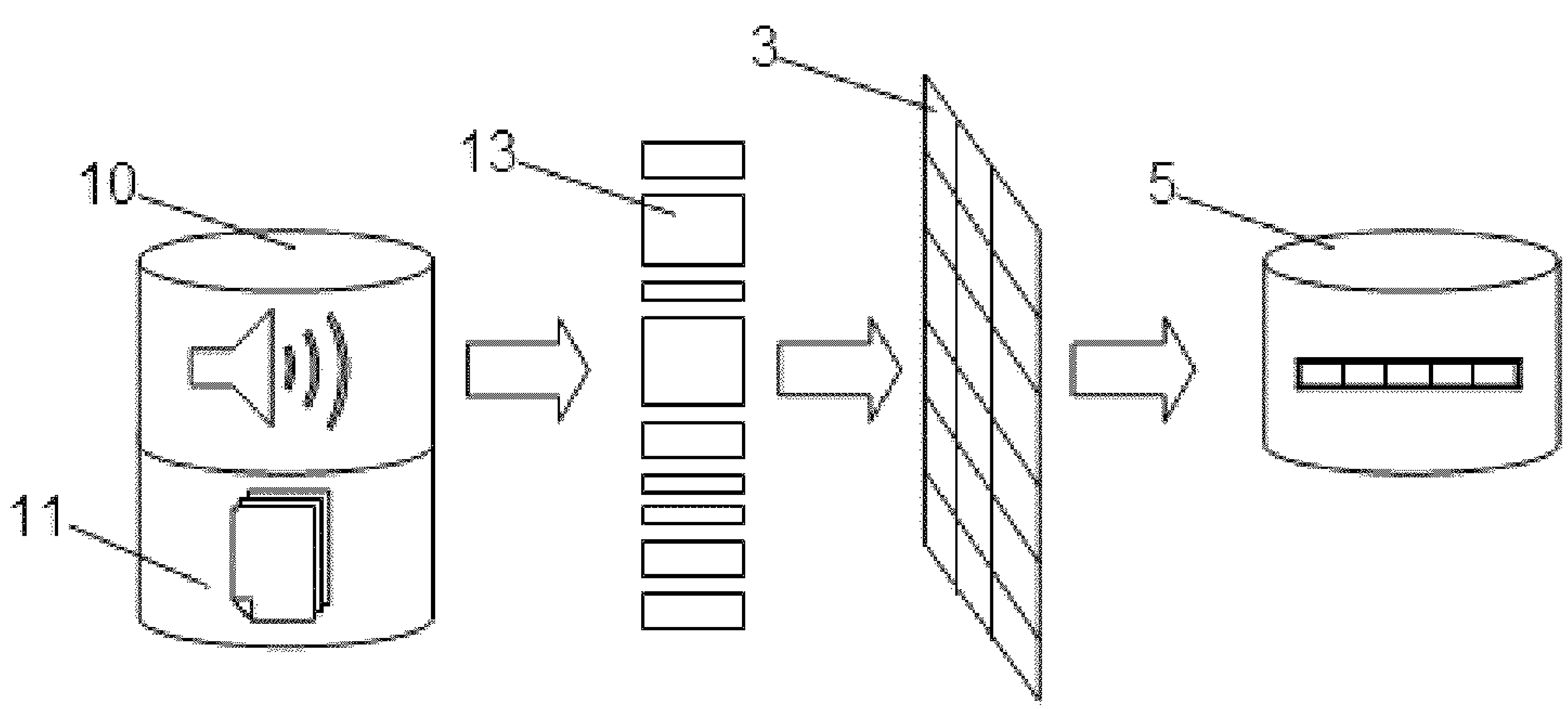

10
11
13
3
5

FIG. 4

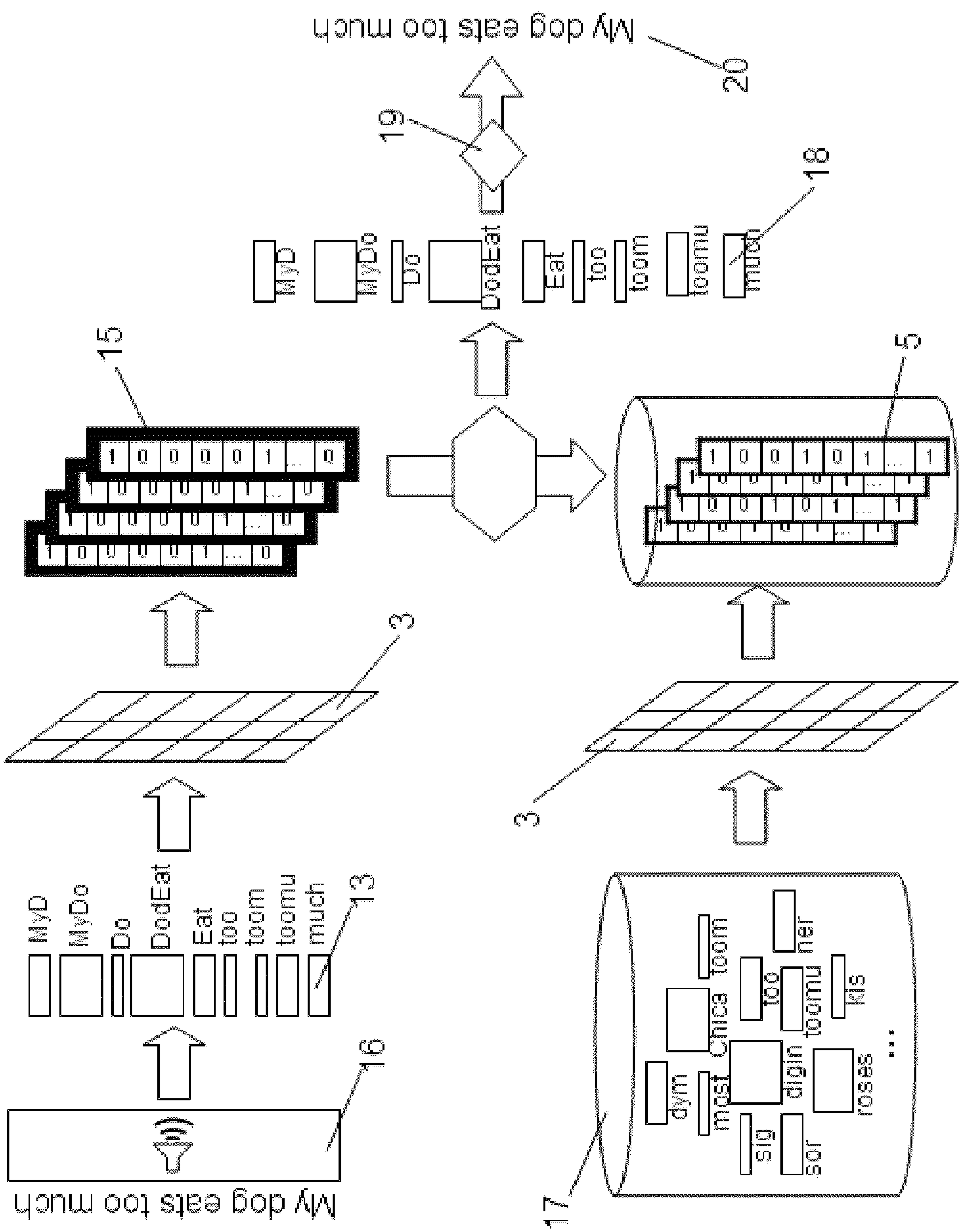

My dog eats too much
16
MyD
MyDo
Do
DodEat
Eat
too
toom
toomu
much
13
3
15
5
3
17
dym
most
Chica
toom
too
toomu
digin
sig
sor
roses
ner
kis
...
MyD
MyDo
Do
DodEat
Eat
too
toom
toomu
much
18
19
20
My dog eats too much

FIG. 5

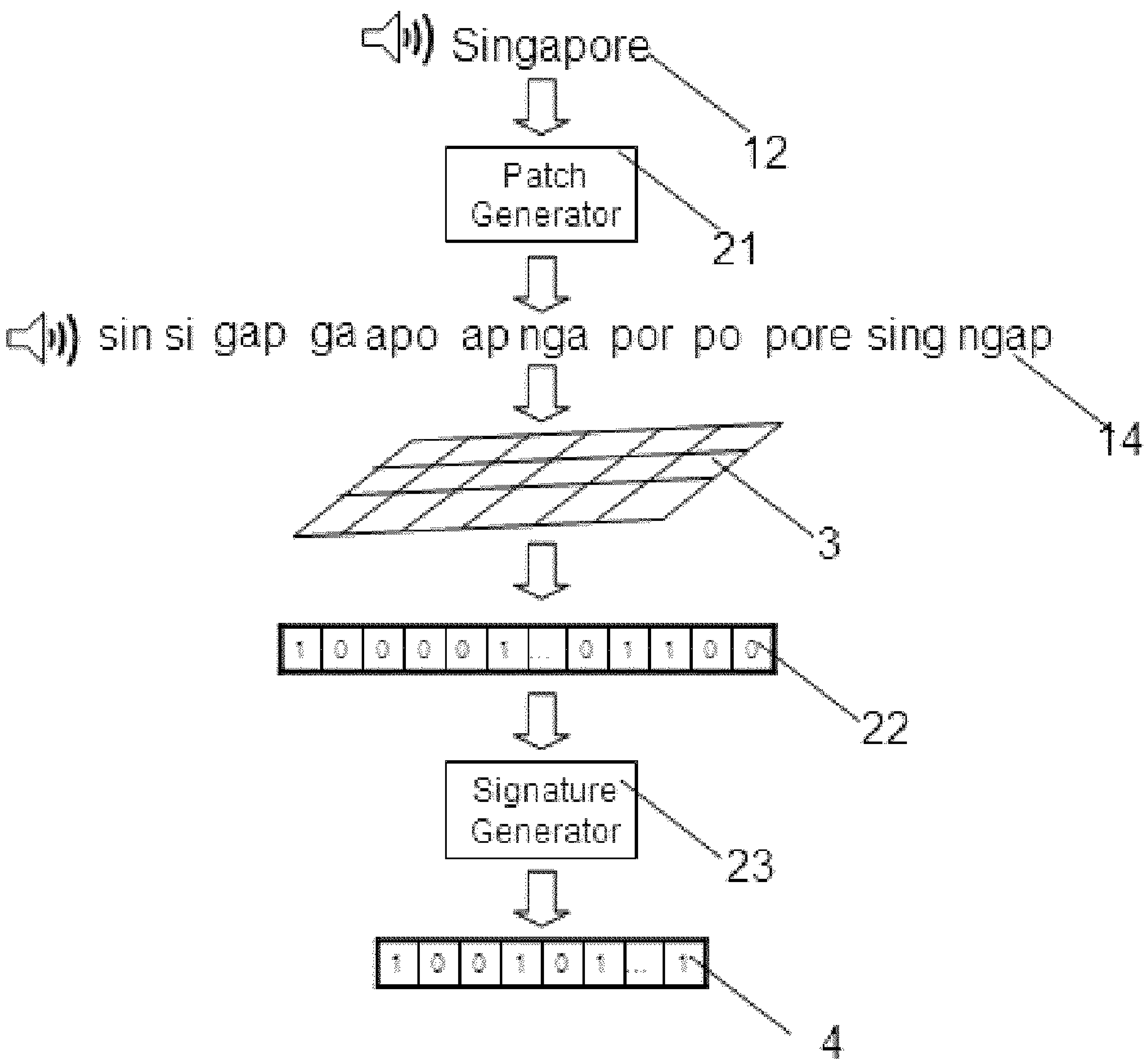

Singapore
12
Patch Generator
21
sin si gap ga apo ap nga por po pore sing ngap
14
3
1 0 0 0 0 1 ... 0 1 1 0 0
22
Signature Generator
23
1 0 0 1 0 1 ... 1
4

FIG. 6

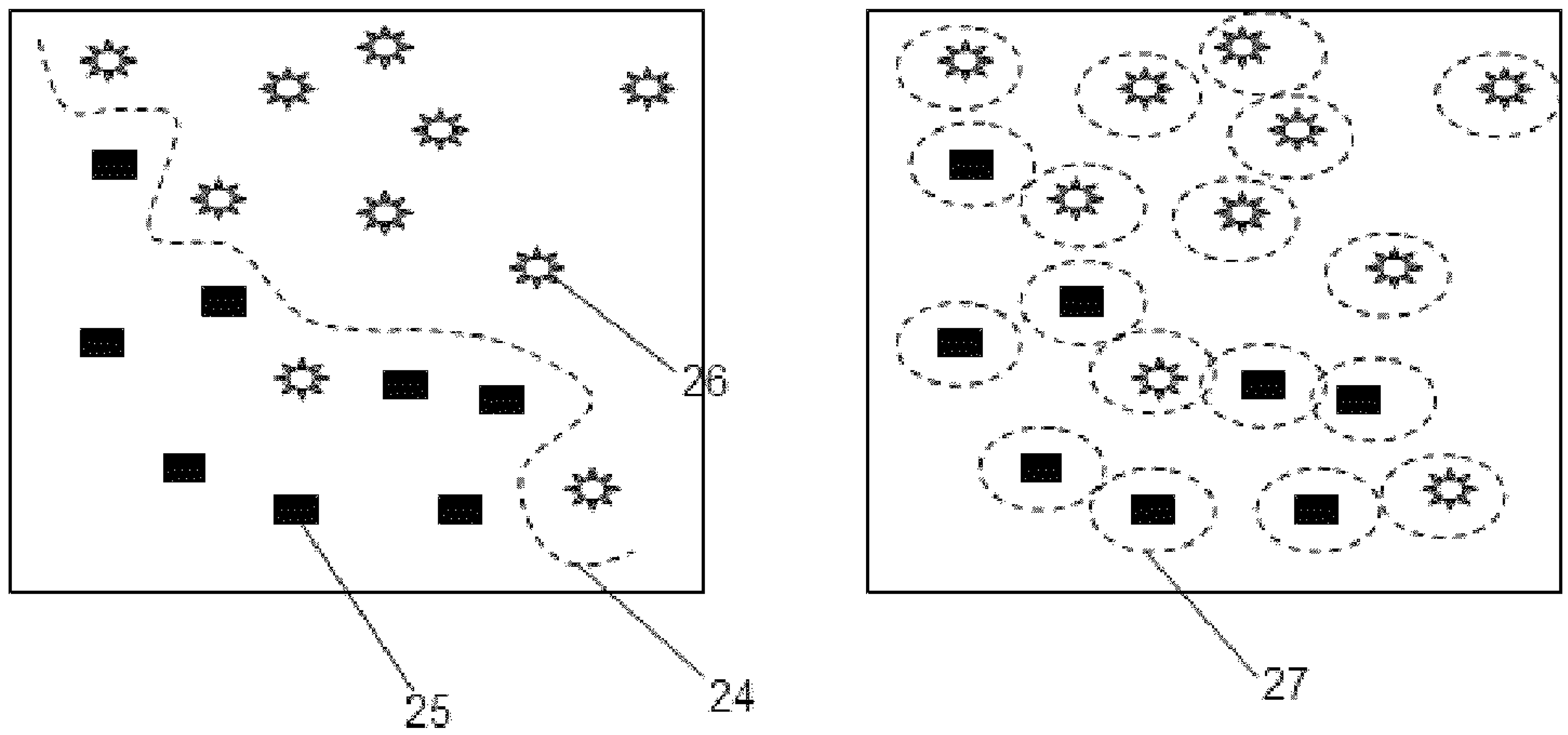


FIG. 7

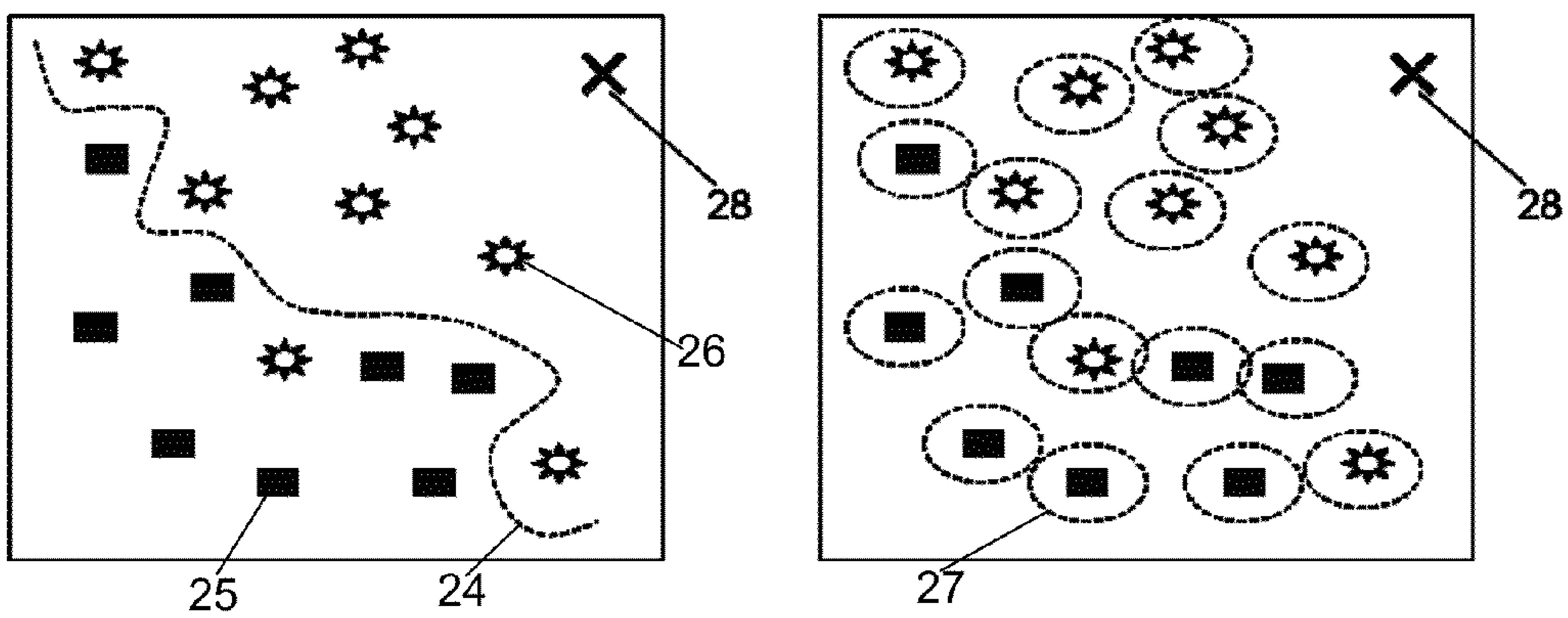


FIG. 8

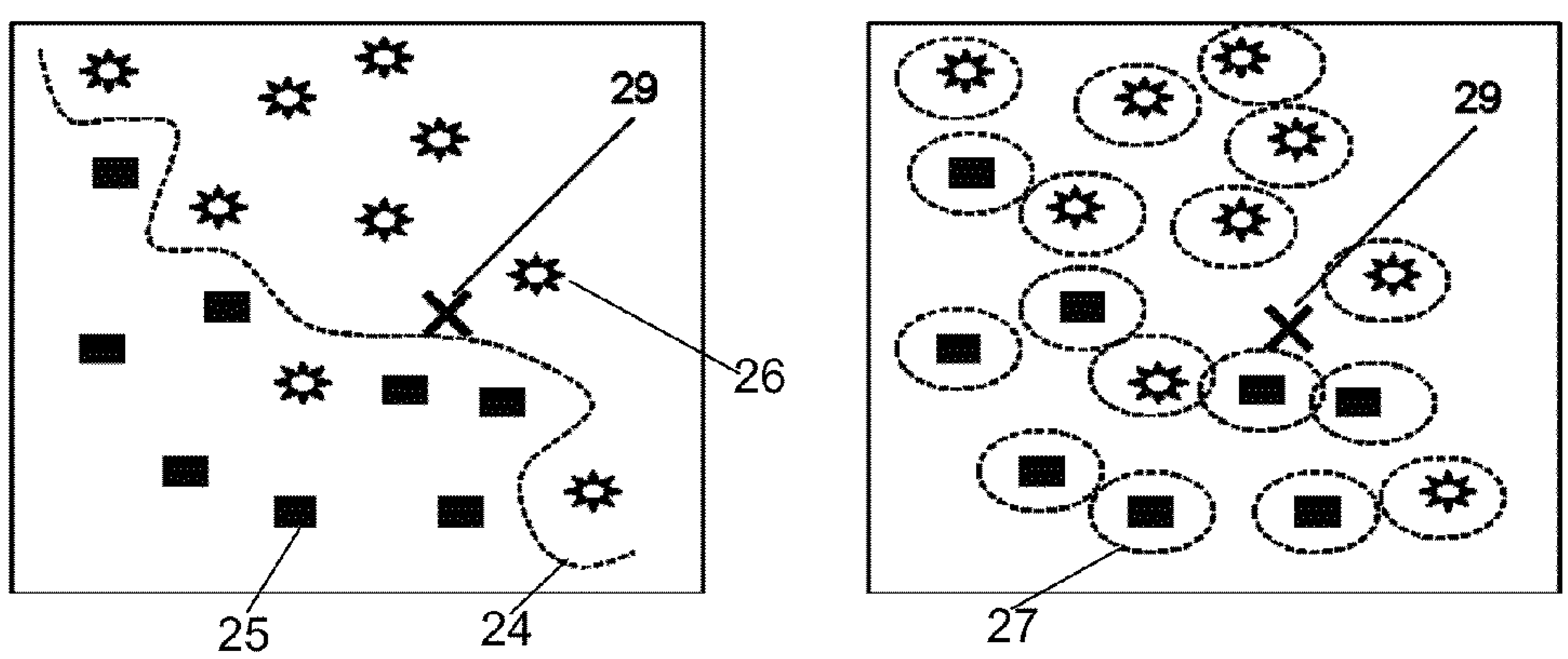


FIG. 9

SIGNATURE GENERATION FOR MULTIMEDIA DEEP-CONTENT-CLASSIFICATION BY A LARGE-SCALE MATCHING SYSTEM AND METHOD THEREOF

CROSS REFERENCE TO RELATED APPLICATIONS

This application is a continuation of U.S. application Ser. No. 12/195,863 filed Aug. 21, 2008, now allowed. This application is also a continuation-in-part of U.S. application Ser. No. 12/084,150 filed on Apr. 7, 2009, now pending, which is the National Stage of International Application No. PCT/IL2006/001235, filed on Oct. 26, 2006, which claims foreign priority from Israeli Application No. 171577 filed on Oct. 26, 2005, and Israeli Application No. 173409 filed on 29 Jan. 2006. This application also claims priority under 35 USC 119 from Israeli Application No. 185414, filed on Aug. 21, 2007, now pending. All of the applications referenced above are herein incorporated by reference.

TECHNICAL FIELD

The invention generally relates to content-based clustering, recognition, classification and search of high volumes of multimedia data in real-time, and more specifically to real-time, fast generation of signatures of high-volume of multimedia content-segments.

BACKGROUND

With the abundance of multimedia data made available through various means in general and the Internet and world-wide web (WWW) in particular, there is also a need to provide for effective ways of searching for such multimedia data. Searching for multimedia data in general and video data in particular may be challenging at best due to the huge amount of information that needs to be checked. Moreover, when it is necessary to find a specific content of video, the prior art cases revert to various metadata that describes the content of the multimedia data. However, such content may be complex by nature and not necessarily adequately documented as metadata.

The rapidly increasing multimedia databases, accessible for example through the Internet, calls for the application of effective means for search-by-content. Searching for multimedia in general and for video data in particular is challenging due to the huge amount of information that has to be classified. Moreover, prior art techniques revert to model-based methods to define and/or describe multimedia data. However, by its very nature, the structure of such multimedia data may be too complex to be adequately represented by means of metadata. The difficulty arises in cases where the target sought for multimedia data cannot be adequately defined in words, or respective metadata of the multimedia data. For example, it may be desirable to locate a car of a particular model in a large database of video clips or segments. In some cases the model of the car would be part of the metadata but in many cases it would not. Moreover, the car may be at angles different from the angles of a specific photograph of the car that is available as a search item. Similarly, if a piece of music, as in a sequence of notes, is to be found, it is not necessarily the case that in all available content the notes are known in their metadata form, or for that matter, the search pattern may just be a brief audio clip.

A system implementing a computational architecture (hereinafter "The Architecture") typically consists of a large ensemble of randomly, independently, generated, heterogeneous processing cores, mapping in parallel data-segments onto a high-dimensional space and generating compact signatures for classes of interest. The Architecture is based on a PCT patent application number WO 2007/049282 and published on May 3, 2007, entitled "A Computing Device, a System and a Method for Parallel Processing of Data Streams", assigned to common assignee, and is hereby incorporated by reference for all the useful information it contains.

It would be advantageous to use The Architecture to overcome the limitations of the prior art described hereinabove. Specifically, it would be advantageous to show a framework, a method, a system and respective technological implementations and embodiments, for large-scale matching-based multimedia deep content classification, that overcomes the well-known limitations of the prior art.

SUMMARY

Certain embodiments disclosed herein include an apparatus for generating a signature for an input signal. The apparatus comprises a processor; and a memory coupled to the processor and configured to store at least instructions causing for execution of a plurality of computational cores, wherein the instructions includes a plurality of subsets of instructions, each subset of instructions causes the execution of one computational core from the plurality of computational cores and is coded to have properties that have at least some statistical independency from other subset of instructions, thereby the properties are independent of each other computational core, wherein each subset of instructions causes its respective computational core to generate responsive of the input signal a first signature element and a second signature element, the first signature element is a robust signature; wherein a plurality of first signature elements include a first signature of the input signal and a plurality of second signature elements include a second signature of the input signal, the first signature is determined respective of a first threshold value and the second signature is determined respective of a second threshold, wherein the first threshold is higher than the second threshold.

Certain embodiments disclosed herein also include a method for large-scale classification of multimedia signals. The method comprises generating by a signature generator a set of robust signatures representing samples for at least one class of multimedia signals; for each multimedia signal to be classified performing: generating at least a signature by the signature generator; determining if the multimedia signal matches the at least one of class of multimedia signals based on the signature generated for the multimedia signal and the set of robust signatures representing samples for the at least one of class of multimedia signals; and classifying the multimedia signal to a class of the multimedia signal matching the at least one of class of multimedia signals.

BRIEF DESCRIPTION OF THE DRAWINGS

The subject matter that is regarded as the invention is particularly pointed out and distinctly claimed in the claims at the conclusion of the specification. The foregoing and other objects, features and advantages of the invention will be apparent from the following detailed description taken in conjunction with the accompanying drawings.

FIG. 1 is the block diagram showing the basic flow of a for large-scale video matching system implemented in accordance with certain embodiments of the invention.

FIG. 2 is a bars-plot showing an example of certain distribution of values of coupling node.

FIG. 3 is an example of a Signature and a corresponding Robust Signature for a certain frame.

FIG. 4 is a diagram depicting the process of generating a signature for a segment of speech implemented in accordance with certain embodiments of the invention.

FIG. 5 is a diagram depicting a process executed by a Large-Scale Speech-to-Text System as implemented in accordance with certain embodiments of the invention.

FIG. 6 is a diagram showing the flow of patches generation, response vector generation, and signature generation in a Large-Scale Speech-to-Text System implemented in accordance with certain embodiments of the invention.

FIG. 7 is a diagram showing the difference between complex hyper-plane generated by prior art techniques, and the large-scale classification techniques where multiple robust hyper-plane segments are generated.

FIG. 8 is a diagram showing the difference in decision making using prior art techniques and the disclosed techniques, when the sample to be classified differs from other samples that belong to the training set.

FIG. 9 is a diagram showing the difference in decision making using prior art techniques and the disclosed techniques, in cases where the sample to be classified closely resembles samples that belong to two classes.

DETAILED DESCRIPTION

It is important to note that the embodiments disclosed by the invention are only examples of the many advantageous uses of the innovative teachings herein. In general, statements made in the specification of the present application do not necessarily limit any of the various claimed inventions. Moreover, some statements may apply to some inventive features but not to others. In general, unless otherwise indicated, singular elements may be in plural and vice versa with no loss of generality. In the drawings, like numerals refer to like parts through several views.

Certain embodiments of the invention include a framework, a method, a system and their technological implementations and embodiments, for large-scale matching-based multimedia Deep Content Classification (DCC). The system is based on an implementation of a computational architecture ("The Architecture") based on "A Computing Device, a System and a Method for Parallel Processing of Data Streams" technology, having a PCT patent application number WO 2007/049282 and published on May 3, 2007. The Architecture consists of a large ensemble of randomly, independently, generated, heterogeneous processing computational cores, mapping in parallel data-segments onto a high-dimensional space and generating compact signatures for classes of interest.

In accordance with the principles of the invention, a realization of The Architecture embedded in large-scale matching system ("The System") for multimedia DCC is disclosed. The Architecture receives as an input stream, multimedia content segments, injected in parallel to all computational cores. The computational cores generate compact signatures for the specific content segment, and/or for a certain class of equivalence and interest of content-segments. For large-scale volumes of data, the signatures are stored in a conventional way in a database of size N, allowing match between the generated signatures of a certain content-segment and the signatures in the database, in low-cost, in terms of complexity, i.e. <=O(log N), and response time.

For the purpose of explaining the principles of the invention there are now demonstrated two embodiments: a Large-Scale Video Matching System; and a Large-Scale Speech-to-Text System. However, it is appreciated that other embodiments will be apparent to one of ordinary skill in the art.

Characteristics and advantages of the System include but are not limited to:

- The System is flat and generates signatures at an extremely high throughput rate;
- The System generates robust natural signatures, invariant to various distortions of the signal;
- The System is highly-scalable for high-volume signatures generation;
- The System is highly-scalable for matching against large-volumes of signatures;
- The System generates Robust Signatures for exact-match with low-cost, in terms of complexity and response time;
- The System accuracy is scalable versus the number of computational cores, with no degradation effect on the throughput rate of processing;
- The throughput of The System is scalable with the number of computational threads, and is scalable with the platform for computational cores implementation, such as FPGA, ASIC, etc.; and
- The signatures produced by The System are task-independent, thus the process of classification, recognition and clustering can be done independently from the process of signatures generation, in the superior space of the generated signatures.

Large-Scale Video Matching System

The goal of a large-scale video matching system is effectively to find matches between members of large-scale Master DB of video content-segments and a large-scale Target DB of video content-segments. The match between two video content segments should be invariant to a certain set of statistical distortions performed independently on two relevant content-segments. Moreover, the process of matching between a certain content-segment from Master DB to Target DB consisting of N segments, cannot be done by matching directly the Master content-segment to all N Target content-segments, for large-scale N, since such a complexity of O(N), will lead to non-practical response times. Thus, the representation of content-segments by both Robust Signatures and Signatures is critical application-wise. The System embodies a specific realization of The Architecture for the purpose of Large-Scale Video Matching System.

A high-level description of the process for large-scale video matching is depicted in FIG. **1**. Video content segments (**2**) from Master and Target databases (**6**) and (**1**) are processed in parallel by a large number of independent computational Cores (**3**) that constitute the Architecture. Further details are provides in the cores generator for Large-Scale Video Matching System section below. The independent Cores (**3**) generate a database of Robust Signatures and Signatures (**4**) for Target content-segments (**5**) and a database of Robust Signatures and Signatures (**7**) for Master content-segments (**8**). The process of signature generation is shown in detail in FIG. **6**. Finally, Target Robust Signatures and/or Signatures are effectively matched, by matching algorithm (**9**), to Master Robust Signatures and/or Signatures database to find all matches between the two databases.

To demonstrate an example of signature generation process, it is assumed, merely for the sake of simplicity and without limitation on the generality of the invention, that the signatures are based on a single frame, leading to certain simplification of the computational cores generation. This is further described in the cores generator for Large-Scale Video Matching System section. The system is extensible for signatures generation capturing the dynamics in-between the frames.

Signature Generation

Creation of Signature Robust to Additive Noise

Assuming L computational cores, generated for Large-Scale Video Matching System. A frame i is injected to all the cores. The cores generate two binary response vectors the Signature $\vec{S}$ and Robust Signature $\overrightarrow{RS}$.

For generation of signatures robust to additive noise, such White-Gaussian-Noise, scratch, etc., but not robust to distortions, such as crop, shift and rotation, the core $C_i=\{n_i\}$ may consist of a single (LTU) node or more nodes. The node equations are:

$$V_i = \sum_j w_{ij} k_j$$

where, $n_i=\theta(V_i-Th_x)$ and

θ is a Heaviside step function; w_{ij} is a coupling node unit (CNU) between node i and image component j (for example, grayscale value of a certain pixel j);

k_j is an image component j (for example, grayscale value of a certain pixel j);

Th_x is a constant Threshold value where x is 'S' for Signature and 'RS' for Robust Signature; and

V_i is a coupling node value.

The Threshold Th_x values are set differently for Signature generation and for Robust Signature generation. For example, as shown in FIG. **2**, for a certain distribution of V_i values (for the set of nodes), the thresholds for Signature Th_S and Robust Signature Th_{RS} are set apart, after optimization, according to the following criteria:

I: For

$$V_i > Th_{RS}$$

$$1-p(V>Th_S)=1-(1-\epsilon)^l>>1$$

i.e., given that/nodes (cores) constitute a Robust Signature of a certain image I, the probability that not all of these/ nodes will belong to the Signature of same, but noisy image, ˜ is sufficiently low (according to a system's specified accuracy).

II:

$$p(V_i>Th_{RS})\approx l/L$$

i.e., approximately/out of the total L nodes can be found to generate Robust Signature according to the above definition.

III: Both Robust Signature and Signature are generated for a certain frame i. An example for generating Robust Signature and Signature for a certain frame is provided in FIG. **3**.

Creation of Signatures Robust to Noise and Distortions

Assume L denotes the number of computational cores in the System. Having generated L cores by the core generator that constitute the Large-Scale Video Matching System, a frame i is injected to all the computational cores. The computational cores map the image frame onto two binary response vectors: the Signature $\vec{S}$ and the Robust Signature $\overrightarrow{RS}$.

In order to generate signatures robust to additive noises, such as White-Gaussian-Noise, scratch, etc., and robust to distortions, such as crop, shift and rotation, etc., the core C_i should consist of a group of nodes (LTUs): $C_i=\{n_{im}\}$, where m is the number of nodes in each core i, generated according to certain statistical process, modeling variants of certain set of distortions.

The first step in generation of distortions-invariant signatures is to generate m Signatures and Robust Signatures, based on each of the m nodes in all the L cores, according to the previously-described (above) algorithm. The next step is to determine a subset V of m potential signatures-variants for certain frame i. This is done by defining a certain consistent and robust selection criterion, for example, select top f signature-variants out of m, with highest firing-rate across all L computational cores. The reduced set will be used as Signature and Robust Signature, invariant to distortions which were defined and used in the process of computational cores generation.

Computational Cores Generation

Computational Cores Generation is a process of definition, selection and tuning the Architecture parameters for a certain realization in specific system and application. The process is based on several design considerations, such as:

(a) The cores should be designed so as to obtain maximal independence, i.e., the projection from a signal space should generate a maximal pair-wise distance between any two computational cores' projections in a high-dimensional space.

(b) The computational cores should be optimally designed for the type of signals, i.e. the computational cores should be maximally sensitive to the spatio-temporal structure of the injected signal, for example, and in particular, sensitive to local correlations in time and space.

(c) The computational cores should be optimally designed with regard to invariance to set of signal distortions, of interest in relevant application.

Following is a non-limiting example of core-generator module for large-scale video-matching system is presented. The first step is a generation of L nodes, 1 for each of the L computational cores, following design optimization criteria (a) and (b).

Criterion (a) is implemented by formulating it as a problem of generating L projections, sampling uniformly a D-dimensional hemisphere. This problem cannot be solved analytically for an arbitrary L. However, there are singular solutions, obtained by Neil Sloane for a certain number of points for a given dimension. The definition of core-generator stochastic process is based on this singular solution. Another constraint embedded in this process definition is local distribution of coupling node currents (CNCs) according to design optimization criterions (b), i.e. the sparse connectivity has local characteristics in image space. Other solutions of almost uniform tessellations exist.

The second step is to fulfill design optimization criterion (c), by generating for each of the nodes of the computational cores, M variants, so that the cores will produce signatures robust to specific distortions of interest. This is done by applying to the functions of each node M.

Large-Scale Speech-to-Text System

The goal of large-scale speech-to-text system is to reliably translate fluent prior art technologies are based on model-based approaches, i.e., speech recognition through phonemes recognition and/or word recognition by Hidden-Markov-Models (HMM) and other methods, natural-language-processing techniques, language models and more, the disclosed approach constitutes a paradigm-shift in the speech-recognition domain. The disclosed System for speech-to-text is based on a previously-disclosed computational paradigm-shift, The Architecture.

FIG. **4** shows high-level steps for generating a signature for voice segment implemented in accordance with certain embodiments of the invention. The System receives a large-scale database of speech (**10**) with relevant database of text (**11**) and generates a database of Robust Signatures (**5**) to patches of the speech signals (**13**) provided in the original database.

FIG. **5** shows more detailed overall process of speech-to-text translation implemented in accordance with certain embodiments of the invention. In the process of speech-to-text translation, the system performs first speech-to-speech match, i.e. the system finds M best matches (**18**) between the speech-segment received as an input (**16**), and the N speech-segments provided in the training database (**17**). Similar to the case of visual signal, the match between two speech-segments should be invariant to a certain set of statistical processes performed independently on two relevant speech-segments, such as generation of the speech by different speakers, plurality noisy channels, various intonations, accents and more. Moreover, the process of matching between a certain speech-segment to a database consisting of N segments, cannot be done by matching directly the speech-segment to all N speech-segments, for large-scale N, since such a complexity of O(N), will lead to non-practical response times. Thus, the representation of speech-segments by Robust Signatures is critical application-wise. The System embodies a specific realization of The Architecture for the purpose of Large-Scale Speech-to-Speech System invention and definition. Finally, after matching the speech-segment to M best matches in database, the relevant text attached to the M segments is post-processed (**19**), generating the text (**20**) of the speech-segment provided as an input.

High-level description of the system is further depicted, in FIG. **5**. Speech-segments are processes by computational Cores (**3**), a realization of The Architecture (see cores generator for Large-Scale Speech-to-Text System). The Cores (**3**) generate a database of Signatures (**5**) for a large-scale database of speech-segments (**17**) and Robust Signatures (**15**) for speech-segment presented as an input (**16**). The process of signature generation is described below. Next, Robust Signatures (**15**) and/or Signatures (**5**) are effectively matched to Robust Signatures (**15**) and/or Signatures (**5**) in the database to find all matches between the two, and finally extract all the relevant text to be post-processed and presented as a text output (**20**).

Signatures Generation

The signatures generation process will be described with reference to FIG. **6**. The first step in the process of signatures generation from a given speech-segment is to break-down the speech-segment to K patches (**14**) of random length P and random position within the speech segment (**12**). The break-down is performed by the patch generator component (**21**). The value of K and the other two parameters are determined based on optimization, considering the tradeoff between accuracy rate and the number of fast matches required in the flow process of the System.

In the next step, all the K patches are injected in parallel to all L computational Cores (**3**) to generate K response vectors (**22**).

Having L computational cores, generated by the cores generator for Large-Scale Speech-to-Text System, a patch i is injected to all the computational cores. Processing by the computational cores yields a response vector $\vec{R}$, for example, in the following way:

A computational core C_i consists of a m nodes (LTUs), generated according to

cores-generator: $C_i=\{n_{im}\}$.

$$n_{im}(t) = \theta(V_i(t) - Th)$$

$$V_i(t) = (1 - L)V_i(t-1) + V_{im}$$

$$V_{im} = \sum_j w_{i,jn} k_{i,j}$$

w_{ij} is a CNU between node j (in Core i) and patch component n (for example, MFCC coefficient), and/or between node j and node n in the same core i.

$k_{i,j}$ is a patch component n (for example, MFCC coefficient), and/or node j and node n in the same core i.

θ is a Heaviside step function; and

Th is a constant threshold value of all nodes.

The response vector $\vec{R}$ is the firing rate of all nodes, $\{n_{im}\}$. The Signature (**4**) and the Robust Signature may be generated, for example, similarly as to the case of video content-segment, i.e., $\vec{S}$ by applying the threshold $\overrightarrow{Th_S}$ to $\vec{R}$ and $\overrightarrow{RS}$ by applying the threshold $\overrightarrow{Th_{RS}}$ to $\vec{R}$.

Speech-to-Speech-to-Text Process

Upon completion of the process of speech-to-speech matching, yielding M best matches from the database, the output of the relevant text is obtained by post-processing (**19**) of the attached text to the M records, for example, by finding the common dominator of the M members.

As an example, if the match yielded the following M=10 attached text records:

This dog is fast
This car is parking
Is it barking
This is a dog
It was barking
This is a king
His dog is playing
He is barking
This dog is nothing
This frog is pink

The output text to the provided input speech-segment will be:

. . . this dog is barking . . .

The proposed System for speech-to-text constitutes a major paradigm-shift from existing approaches to the design of prior art speech-to-text systems in several aspects. First, it is not model-based, i.e. no models are generated for phonemes, key-words, speech-context, and/or language. Instead, signatures generated for various speech-fragments, extract this information, which is later, easily retrieved, by low-cost database operation during the recognition process. This yields a major computational advantage in that no expert-knowledge of speech understanding is required during the training process, which in the disclosed method and its embodiment is signature generation. Second, the System does not require an inference of the input speech-segment to each of the generated models. Instead, for example, the Robust Signature generated for the input segment is matched against the whole database of signatures, in a way which does not require a complexity greater than O(log N). This yields inherent scalability characteristic of the System, and extremely short response times.

Synthesis for Generation of Large-Scale "Knowledge" Databases

One of the main challenges in developing speech-to-text systems, with superior performance, is the process of collecting a large-scale and heterogeneous enough, "training" database. In the sequel, an innovative approach for meeting this challenge is presented. For the purpose of large-scale database generation of transcribed speech, a prior art synthesizers is used. A synthesizer receives two inputs: (1) Large text database (2) Speech data-base with multiple speakers, intonations, etc., and generates a large database of heterogeneous speech, transcribed according to the provided text database. The generated large-scale database of transcribed speech is used according to the presented System flow.

Large-Scale Classification Paradigm-Shift

The presented System implements a computational paradigm-shift required for classification tasks of natural signals, such as video and speech, at very large scales of volume and speed. For very large-scale tasks, such as the classification tasks related to the web content and/or any other large-scale database in terms of volume and update frequencies, the required performance envelope is extremely challenging. For example, the throughput rate of The System signature generation process should be equal to the rate of update process of the content database. Another example is the false-alarm or false-positive rate required for the System to be effective. 1% false-positive rate for a certain content-segment may turn to 100% false-positive rate for a data-base of N content-segments being matched against another large-scale data-base. Thus, the false-positive rates should be extremely low. The presented System does afford such a low false-positive rate due to the paradigm-shift in its computational method for large-scale classification tasks. Unlike, prior art learning systems which generate a complex hyper-plane separating a certain class from the entire "world", and/or model-based method, which generate a model of a certain class, the presented System generates a set of Robust Signatures for the presented samples of the class according to teachings described above. Specifically, the signatures are generated by maximally independent, transform/distortions-invariant and signal-based characteristics optimally designed computational cores. The generalization from a certain set of samples to a class is well defined in terms of invariance to transforms/distortions of interest, and the signatures' robustness, yielding extremely low false-positive rates. Moreover, the accuracy is scalable by the signatures length due the low dependence of the computational cores.

Several differences between the prior art techniques and the scale classification techniques disclosed herein are illustrated in FIGS. **7**, **8** and **9**. Specifically, FIG. **7** shows a diagram illustrating the difference between complex hyper-plane the large-scale classification where multiple robust hyper-plane segments and are generated, prior art classification shown on the left and classification according to the principles of the invention on the right. Prior art classification attempts to find a sophisticated classification line (**24**) that best separates between objects (**25**) belonging to one group and objects (**26**) that belong to another group. Typically, one or more of the objects of one group are found to be classified into the other group, in this example, there is an object (**26**) within the group of different objects (**25**). In accordance with an embodiment of the invention, each object is classified separately (**27**) and matched to its respective objects. Therefore an object will belong to one group or another providing for a robust classification. FIG. **8** illustrates the difference in decision making when the sample to be classified differs from other samples that belong to the training set, prior art classification shown on the left and classification according to the principles of the invention on the right. When a new object (**28**), not previously classified by the system is classified according to prior art as belonging to one group of objects, in this exemplary case, objects (**26**). In accordance with the disclosed invention, as the new object (**28**) does not match any object (**27**) it will be recorded as unrecognized, or no match. FIG. **9** shows the difference in decision making in cases where the sample to be classified closely resembles samples that belong to two classes, prior art classification shown on the left and classification according to the principles of the invention on the right. In this case the new object (**29**) is classified by prior art systems as belonging to one of the two existing, even though line (**24**) may require complex computing due to the similarity of the new object (**29**) to wither one of the objects (**25**) and (**26**). However, in accordance with an embodiment of the invention, as each object is classified separately (**27**) it is found that the new object (**29**) does not belong to any one of the previously identified objects and therefore no match is found.

The foregoing detailed description has set forth a few of the many forms that the invention can take. It is intended that the foregoing detailed description be understood as an illustration of selected forms that the invention can take and not as a limitation to the definition of the invention. It is only the claims, including all equivalents that are intended to define the scope of this invention.

Most preferably, the principles of the invention are implemented as any combination of hardware, firmware and software. Moreover, the software is preferably implemented as an application program tangibly embodied on a program storage unit or computer readable medium. The application program may be uploaded to, and executed by, a machine comprising any suitable architecture. Preferably, the machine is implemented on a computer platform having hardware such as one or more central processing units ("CPUs"), a memory, and input/output interfaces. The computer platform may also include an operating system and microinstruction code. The various processes and functions described herein may be either part of the microinstruction code or part of the application program, or any combination thereof, which may be executed by a CPU, whether or not such computer or processor is explicitly shown. In addition, various other peripheral units may be connected to the computer platform such as an additional data storage unit and a printing unit.

What we claim is:

1. An apparatus for generating a signature for an input signal, comprising:

a processor; and

a memory coupled to the processor and configured to store at least instructions causing for execution of a plurality of computational cores, wherein the instructions includes a plurality of subsets of instructions, each subset of instructions causes the execution of one computational core from the plurality of computational cores and is coded to have properties that have at least some statistical independency from other subset of instructions, thereby the properties are independent of each other computational core, wherein each subset of instructions causes its respective computational core to generate responsive of the input signal a first signature element and a second signature element, the first signature element is a robust signature;

wherein a plurality of first signature elements include a first signature of the input signal and a plurality of second signature elements include a second signature of the input signal, the first signature is determined respective of a first threshold value and the second signature is determined respective of a second threshold, wherein the first threshold is higher than the second threshold.

2. The apparatus of claim **1**, wherein the first signature is robust to at least one of: noise and distortion.

3. The apparatus of claim **1**, wherein the properties include in part a first threshold and a second threshold, each of the plurality of computational cores generates responsive to the input signal and the first threshold the first signature element and responsive to the input signal and the second threshold the second signature element.

4. The apparatus of claim **3**, wherein each of the first threshold and the second threshold is a constant threshold value of a Heaviside step function, wherein each of the subset of instructions are coded with the Heaviside step function to be computed by each respective computational core when generating the first signature element and the second signature element.

5. The apparatus of claim **3**, wherein the properties are defined according to at least one random parameter.

6. The apparatus of claim **1**, wherein the input signal is at least multimedia signal.

7. The apparatus of claim **6**, wherein the multimedia signal is at least one of: an image, graphics, a video stream, a video clip, an audio stream, an audio clip, a video frame, text, a photograph, images of signals, and portions thereof.

8. A method for large-scale classification of multimedia signals, comprising:

generating by a signature generator a set of robust signatures representing samples for at least one class of multimedia signals;

for each multimedia signal to be classified performing:

generating at least a signature by the signature generator;

determining if the multimedia signal matches the at least one of class of multimedia signals based on the signature generated for the multimedia signal and the set of robust signatures representing samples for the at least one of class of multimedia signals; and

classifying the multimedia signal to a class of the multimedia signal matching the at least one of class of multimedia signals.

9. The method of claim **8**, further comprising:

creating a new class of multimedia signals to include an unmatched multimedia signal if is a match does not exist.

10. The method of claim **8**, wherein each robust signature of the set of robust signatures is robust to noise and distortion.

11. The method of claim **8**, wherein robust signatures of the set of robust signatures and the signature of the multimedia signal are generated by a signature generator.

12. The method of claim **11**, wherein the signature generator includes:

a processor; and

a memory coupled to the processor and configured to store at least instructions causing for execution of a plurality of computational cores, wherein the instructions includes a plurality of subsets of instructions, each subset of instructions causes the execution of one computational core from the plurality of computational cores and is coded to have properties that have at least some statistical independency from other subset of instructions, thereby the properties are independent of each other computational core, wherein each subset of instructions causes its respective computational core to generate responsive of a multimedia signal a first signature element and a second signature element, the first signature element is a robust signature;

wherein a plurality of first signature elements include a first signature of the input signal and a plurality of second signature elements include a second signature of the input signal, the first signature is determined respective of a first threshold value and the second signature is determined respective of a second threshold, wherein the first threshold is higher than the second threshold.

13. The method of claim **8**, wherein the multimedia signal is at least one of: an image, graphics, a video stream, a video clip, an audio stream, an audio clip, a video frame, text, a photograph, images of signals, and portions thereof.

* * * * *

UNITED STATES PATENT AND TRADEMARK OFFICE

CERTIFICATE OF CORRECTION

PATENT NO. : 8,990,125 B2
APPLICATION NO. : 13/682132
DATED : March 24, 2015
INVENTOR(S) : Igal Raichelgauz, Karina Odinaev and Yehoshua Y. Zeevi

It is certified that error appears in the above-identified patent and that said Letters Patent is hereby corrected as shown below:

Title page, column 1, item (72):

Inventors: Igal Raichelgauz, Herzelia (IL); Karina “Ordinaev” should read --Odinaev--

Signed and Sealed this
First Day of September, 2015

Michelle K. Lee
Director of the United States Patent and Trademark Office